Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families

Emily J. Capra,1 Barrett S. Perchuk,1 Jeffrey M. Skerker,3 and Michael T. Laub1,2,*
1Department of Biology
2Howard Hughes Medical Institute
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
3Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA
*Correspondence: laub@mit.edu
http://dx.doi.org/10.1016/j.cell.2012.05.033

SUMMARY

Orthologous proteins often harbor numerous substitutions, but whether these differences result from neutral or adaptive processes is usually unclear. To tackle this challenge, we examined the divergent evolution of a model bacterial signaling pathway comprising the kinase PhoR and its cognate substrate PhoB. We show that the specificity-determining residues of these proteins are typically under purifying selection but have, in α-proteobacteria, undergone a burst of diversification followed by extended stasis. By reversing mutations that accumulated in an α-proteobacterial PhoR, we demonstrate that these substitutions were adaptive, enabling PhoR to avoid crosstalk with a paralogous pathway that arose specifically in α-proteobacteria. Our findings demonstrate that duplication and the subsequent need to avoid crosstalk strongly influence signaling protein evolution. These results provide a concrete example of how system-wide insulation can be achieved postduplication through a surprisingly limited number of mutations. Our work may help explain the apparent ease with which paralogous protein families expanded in all organisms.

INTRODUCTION

The evolutionary forces and selective pressures that influence protein sequences remain poorly understood at a detailed, molecular level. A comparison of orthologs often reveals tens to hundreds of amino acid differences. How and why do functionally equivalent proteins diverge in different organisms? Many of the accumulated substitutions may be functionally neutral and result from processes such as genetic drift. However, some mutations may have been adaptive and provided a fitness advantage. Identifying these beneficial mutations and pinpointing the advantage that they provide are difficult problems. Comparative sequence analyses, such as measures of codon substitution patterns or dN/dS ratios (Yang and Bielawski, 2000), can help to identify residues that are potentially adaptive, but such approaches are frequently insufficient and difficult to validate. Additionally, elucidating why certain mutations are beneficial requires a genetically manipulatable organism and an ability to probe the effects of individual mutations in vivo.

In many cases where protein evolution has been studied experimentally (reviewed in Dean and Thornton, 2007), the relevant proteins were examined in vitro or in heterologous hosts, and thus outside their native cellular context, possibly eliminating or obscuring important evolutionary constraints. For example, signal transduction proteins are often part of large paralogous families that expand through duplication and divergence. The duplication-divergence process thus runs an inherent risk of introducing crosstalk with existing pathways. A study of SH3 domains from Saccharomyces cerevisiae and humans suggested that the avoidance of crosstalk may represent an important selective pressure in the evolution of paralogous protein families (Zarrinpar et al., 2003). However, a direct demonstration that crosstalk influences the evolution of signaling proteins and, more importantly, an understanding of how this occurs at the amino acid level are lacking.

To tackle these challenges, we examined the evolution of two-component signal transduction proteins in bacteria. These pathways, a primary means of signal transduction in prokaryotes, typically involve a sensor histidine kinase that, upon receipt of an input stimulus, autophosphorylates and then transfers its phosphoryl group to a cognate response regulator, which in turn modulates gene expression (Stock et al., 2000). Most histidine kinases are bifunctional and can, in the absence of an input signal, stimulate the dephosphorylation of their cognate response regulators, effectively acting as phosphatases (Huynh and Stewart, 2011).

Although most bacteria encode between 20 and 200 two-component signaling pathways (Alm et al., 2006), very little crosstalk occurs at the level of phosphotransfer in vivo (Grimsshaw et al., 1998; Laub and Goulian, 2007; Siryaporn and Goulian, 2008; Skerker et al., 2005). Two-component pathways are highly specific, typically with one-to-one relationships between
cognate kinases and regulators. When not stimulated to auto-
phosphorylate, the phosphatase activity of a given histidine
kinase can help to eliminate crosstalk and the errant phosphor-
ylation of its cognate regulator (Siryaporn and Goulian, 2008).
However, when stimulated as a kinase, molecular recognition
is the dominant mechanism for preventing phosphotransfer
crosstalk and thereby maintaining the fidelity of distinct sig-
naling pathways. Systematic analyses of phosphotransfer
have demonstrated that histidine kinases are endowed with
an intrinsic ability to discriminate their in vivo cognate substrate
from all other noncognate substrates (Skerker et al., 2005).
Analyses of amino acid coevolution in cognate signaling
proteins identified the key specificity-determining residues in
histidine kinases and response regulators (Bell et al., 2010;
Capra et al., 2010; Casino et al., 2009; Skerker et al., 2008;
Weigt et al., 2009). Rational mutagenesis of these residues
can reprogram the partnering specificity of a histidine kinase
or response regulator (Bell et al., 2010; Capra et al., 2010;
Skerker et al., 2008).

For a given two-component pathway, there is likely strong
purifying selective pressure on its key specificity-determining
residues to preserve the kinase-substrate interaction. Even
single amino acid changes in specificity residues can drastically
change the interaction capabilities and preferences of a histidine
kinase or response regulator (Capra et al., 2010). Nevertheless,
an inspection of orthologous kinases or response regulators
often reveals divergent evolution and variability in the specificity
residues of certain subsets of orthologs, raising the question of
whether these changes resulted from neutral or adaptive pro-
cesses. We favored the latter, hypothesizing that specificity resi-
dues must change in order to avoid crosstalk between pathways
following gene duplication events. Two-component signaling
pathways often expand through duplication (Ahn et al., 2006),
and following such events, bacteria presumably must accumu-
late mutations that insulate the new pathways from each other
and maintain their isolation from other, existing two-component
pathways. Here, we provide direct experimental evidence that
the avoidance of crosstalk is indeed a major selective force in
the evolution of two-component signaling pathways. Through
in vitro studies and fitness competition assays, we identify
specific substitutions in a model two-component pathway,
PhoR-PhoB, that represent an adaptation to the duplication of
another two-component signaling system. Similar adaptations
likely accompanied each of the duplication and divergence
events underlying the massive expansion of two-component
signaling protein families in bacteria. Accordingly, global anal-
yses of specificity-determining residues in extant bacterial
genomes reveal a pervasive trend toward orthogonality in these
signaling proteins.

RESULTS

Vertical Inheritance of PhoR and PhoB
To examine the divergent evolution of two-component signaling
pathways, we focused on the PhoR-PhoB signaling pathway
(Wanner and Chang, 1987), which is found throughout the bacte-
rial kingdom and helps a wide range of organisms respond to
phosphate starvation. To systematically identify orthologs of

Escherichia coli phoR and phoB, we used a modified version
of reciprocal best hits in BLAST analysis that allows for the iden-
tification of putative duplications. Most proteobacteria, except
a small number of α-proteobacteria, were found to encode
a single ortholog of each phoR and phoB, suggesting that these
genes have rarely been duplicated, particularly in the α-, β-, and
γ-proteobacteria (Figure 1A). Additionally, gene trees for phoR
and phoB closely matched a species tree (Figure 1B; Figures
S1A and S1B available online), indicating that this signaling
system has likely been vertically inherited in these clades.

Given that phoR and phoB genes were rarely duplicated
during the evolution of proteobacteria, it might be expected
that the residues dictating phosphotransfer specificity would
be relatively constant in order to preserve the interaction
between PhoR and PhoB. We thus examined the six residues
in PhoR and seven residues in PhoB previously identified as crit-
ical determinants of specificity in two-component signaling
proteins (Capra et al., 2010). We extracted these residues, here-
after referred to simply as specificity residues, from 149 PhoR
orthologs and 92 PhoB orthologs, and built sequence logos
representing the relative frequency of amino acids at each spec-
ificity position (Figure 1B; Table S1). The difference in number
of PhoR and PhoB orthologs results from the independent identifi-

cation of kinase and regulator orthologs; most organisms
encode both PhoR and PhoB.

The specificity residues of both PhoR and PhoB are generally
well conserved (Figures 1B and S1C), although several positions
showed substantial variability. We then split the PhoR and PhoB
sequences into groups corresponding to the three proteobacte-
rial subdivisions, α, β, and γ. Sequence logos built for each
phylogenetic group revealed that differences between subdivi-
sions can account for nearly all of the variability in the combined
sequence logos (Figure 1B). For instance, in γ- and β-proteobac-
teria the first two positions are almost always threonine and
valine, whereas in α-proteobacteria these positions are usually
alanine and serine or two alanines. Similar observations were
made for the specificity residues of PhoB orthologs grouped
according to phylogenetic subdivision. Importantly, each PhoR
and PhoB sequence logo was built using species that are highly
diverged. The strong conservation within each clade thus sugg-
ests that specificity residues are usually subject to strong puri-
fying selection. Why, though, have specificity residues diverged
between clades?

Identification of Adaptive Mutations that Prevent
Crosstalk In Vitro

The clade-specific differences in PhoR and PhoB specificity resi-
dues may simply reflect degeneracy in the residues that enable
PhoR and PhoB to interact. Alternatively, the differences may
have produced functional changes such that a PhoR from one
clade is less efficient at interacting with a PhoB from a different
clade. To distinguish between these possibilities, we purified
PhoR kinases from representative γ- and α-proteobacteria,
E. coli and Caulobacter crescentus, and examined their ability
to phosphorylate a panel of 11 PhoB orthologs from α, β, and
γ-proteobacteria (Figures 1C and S1D). For each PhoB from
a γ-proteobacterium, phosphotransfer from the E. coli (γ) PhoR
was significantly faster than from C. crescentus (α) PhoR.
Similarly, each PhoB from an α-proteobacterium was preferentially phosphorylated by the α-PhoR. For the two chosen β-PhoR orthologs, we observed more rapid phosphorylation by the γ-PhoR than the α-PhoR, consistent with the specificity residues of the β-PhoR and β-PhoB orthologs being more similar to those found in γ-proteobacteria than those in α-proteobacteria. We conclude that within each proteobacterial clade, the phosphotransfer specificity of PhoR and PhoB orthologs is relatively static. However, substitutions in the specificity residues of α-PhoR and α-PhoB orthologs have led to significant differences in phosphotransfer specificity between clades.

The changes in PhoR-PhoB specificity residues, and consequent alteration of interaction specificity could have resulted from neutral drift. However, the strong conservation of specificity residues within each clade, which includes species that are widely divergent, suggests that such drift is extremely rare. Instead, the alternative PhoR-PhoB specificity residues in α-proteobacteria may be adaptive and provide an important selective advantage. We hypothesized that the substitutions in α-PhoR and α-PhoB specificity residues prevent unwanted crosstalk with another pathway that is specific to the α-proteobacteria, i.e., negative selection led to changes in α-PhoR and α-PhoB. This model predicts that PhoR orthologs from γ-proteobacteria may phosphorylate response regulators found exclusively in α-proteobacteria, which the α-PhoR orthologs have adapted to avoid phosphorylating.

To test this possibility, we performed comprehensive phosphotransfer profiling of E. coli (γ) and C. crescentus (α) PhoR. Both PhoR constructs were autophosphorylated in vitro and then examined, in parallel, for phosphotransfer to the 44 response regulators encoded by C. crescentus (Figure 2). Both PhoR constructs phosphorylated the C. crescentus PhoB, consistent with their orthologous relationship; although as noted above, phosphotransfer from the γ-PhoR is more robust. Interestingly, the γ-PhoR showed significant phosphotransfer to NtrX, whereas the α-PhoR construct did not. Notably, most
α-proteobacteria encode two paralogous Ntr systems, NtrB-NtrC and NtrX-NtrY, while the γ-proteobacteria typically encode only one, NtrB-NtrC (Figure 3A; Table S1). The two α-Ntr systems, which likely arose through duplication and divergence, do not crosstalk with each other in vitro (Figure 3B) and, consistently, have different specificity residues (Figure 3C). Collectively,
our observations suggest that the different PhoR specificity residues seen in \( \alpha \)-proteobacteria may have evolved to accommodate the presence of a second, lineage-specific pathway, NtrX-NtrY. Such a change in PhoR was presumably accompanied by changes in the PhoB specificity residues (see Figure 1B) to maintain phosphotransfer from PhoR.

Thus, we hypothesized that the alanine, serine, and phenylalanine found at specificity positions 1, 2, and 4 of \( \alpha \)-PhoR proteins represent adaptive mutations that prevent crosstalk to NtrX. To test this hypothesis, we created a series of \( C. \text{crescentus} \) PhoR mutants in which specificity residues were replaced with the corresponding residues from \( \gamma \)-proteobacterial PhoR. We made each single mutant, three double mutants, and the triple mutant. Each mutant kinase was then profiled against the complete set of \( C. \text{crescentus} \) response regulators to examine what effect, if any, these residues have on phosphotransfer specificity. Strikingly, each mutant led to a significant increase in NtrX phosphorylation (Figure 2).

We also examined detailed time courses of phosphotransfer from each mutant PhoR, as well as the wild-type kinases, to the \( C. \text{crescentus} \) regulators PhoB and NtrX. Each mutant kinase exhibited an increase in crosstalk with NtrX compared to the wild-type \( C. \text{crescentus} \) PhoR, but retained the ability to phosphorylate \( C. \text{crescentus} \) PhoB at rates comparable to the wild-type PhoR (Figures 3D and S2A–S2C). Although some mutant PhoR kinases phosphorylated several substrates (see Figure 2), we focused on PhoB and NtrX as time courses of phosphotransfer indicated these as the two preferred targets of mutant PhoR constructs (Figures S2D and S2E).

The most significant crosstalk to NtrX occurred for PhoR mutants with a valine substituted for serine at specificity position 2. Importantly, substantial crosstalk was not observed when this serine was substituted with other residues including leucine, aspartate, glutamate, and threonine. Only valine, corresponding to that found in \( \gamma \)-proteobacterial PhoR orthologs, produced significant crosstalk (Figures S2F and S2G).

Taken together, our in vitro studies support the notion that alanine, serine, and phenylalanine at specificity positions 1, 2, and 4 represent adaptive mutations that prevent crosstalk to NtrX in \( \alpha \)-proteobacteria.

**Avoidance of Crosstalk Is a Significant Selective Pressure**

To test whether these mutations also prevent crosstalk in vivo, we engineered the chromosomal copy of \textit{phoR} in the \( \alpha \)-proteobacterium \textit{C. crescentus} to produce a mutant PhoR in which specificity positions 1 and 2 are threonine and valine, respectively, as they are in \( \gamma \)-proteobacteria; hereafter, this mutant strain is referred to as PhoR(TV). Based on our in vitro experiments, we expected that crosstalk from PhoR(TV) to NtrX would be induced during growth in phosphate-limited media (Figure 4A). During growth in such conditions, wild-type PhoR is stimulated to autophosphorylate and phosphotransfer to PhoB, which then activates genes involved in responding to...
phosphate limitation. Thus, any effects of increased crosstalk to NtrX by the PhoR(TV) kinase should be manifest specifically during growth in phosphate-limited media.

We grew cells to midlogarithmic phase in phosphate-limited media and measured the rate of growth by monitoring the accumulation of optical density at 600 nm. In minimal media containing either 50 μM phosphate or 5 μM phosphate, the PhoR(TV) mutant grew significantly more slowly than wild-type, with a doubling time 30% longer than wild-type in each case (Figures 4B, S3A, and S3B). This growth defect was almost as severe as that observed for a ΔphoR strain which cannot mount a proper transcriptional response to phosphate limitation. To assess whether crosstalk from PhoR(TV) to NtrX contributed to the slow growth phenotype observed, we deleted ntrX in the PhoR(TV) strain. Indeed, the deletion of ntrX significantly reduced the growth deficiency of the PhoR(TV) mutant (Figures 4B, S3A, and S3B) suggesting that crosstalk with NtrX contributes significantly to the slow growth phenotype of a PhoR(TV) strain. The suppression observed was not a nonspecific acceleration of growth as the ntrX deletion alone had no effect on growth in phosphate-limited medium.

In phosphate-replete medium, the PhoR(TV) mutant strains grew at a rate nearly identical to the wild-type (Figure 4B), indicating that, as expected, crosstalk to NtrX requires PhoR to be activated as a kinase. The ntrX deletion and PhoR(TV)/ΔntrX strains grew more slowly in phosphate-replete medium, as the NtrY-NtrX pathway is likely necessary for responding to a signal or metabolite produced in M2G medium.

To corroborate our growth rate measurements, we performed competitive fitness assays in which each mutant strain was mixed with the wild-type at a ratio of 1:1 and grown in the same flask for 10^4 hr, or approximately 40 wild-type generations.
The mutant and wild-type strains were engineered to constitutively produce cyan fluorescent protein (CFP) or yellow fluorescent protein (YFP), allowing for a rapid assessment of relative strain abundance using fluorescence microscopy. In phosphate-limited conditions, the PhoR(TV) strain showed a significant growth disadvantage, being almost completely eliminated from the population after 104 hr (Figures 4C and S3C). The fitness disadvantage of the PhoR(TV) mutant was comparable to that of ΔphoR competed against wild-type in the same phosphate-limited medium. Consistent with our growth measurements, deleting ntrX in the PhoR(TV) background improved competitive fitness (Figures 4C, S3C, and S3D). In phosphate-replete conditions, the PhoR(TV) and ΔphoR mutants retained a ratio with wild-type close to 1:1, demonstrating that the selective disadvantage of introducing ancestral specificity residues into PhoR likely occurs only in conditions in which PhoR is a kinase. Collectively, these data further support a model in which the α-specific substitutions in PhoR specificity residues (T→A and V→S at specificity positions 1 and 2) are selectively advantageous because they help prevent phosphotransfer crosstalk to NtrX, and perhaps other response regulators.

The growth and competitive fitness defects of PhoR(TV) in phosphate-limited media were comparable to that seen for ΔphoR. This similarity suggested that the detrimental effect of crosstalk in the PhoR(TV) strain stems from an inability to phosphorylate PhoB and activate PhoB-dependent genes in phosphate-limited conditions. To test this hypothesis directly, we examined global gene expression patterns in the PhoR(TV) and ΔphoR strains during growth in phosphate-limited conditions. These expression profiles exhibited strong similarity with a Pearson correlation coefficient of ~0.9 (Table S2), supporting a model in which phosphorylation crosstalk from PhoR(TV) to NtrX specificity residues into an ΔNtrX may eliminate crosstalk from a ΔNtrX specificity residues into an ΔNtrX in vitro and in vivo. Indeed, whereas C. crescentus NtrX was robustly phosphorylated by PhoR(TV), a mutant NtrX harboring the β-like substitutions D13G, G20E, F107I, and K108A was not detectably phosphorylated (Figure 4E). This mutant NtrX was not simply unfolded or unphosphorylatable as it was still robustly phosphorylated by α-NtrY. Hence, the substitutions introduced specifically eliminated crosstalk from PhoR(TV), while still allowing for interaction with the cognate kinase NtrY. Taken together, these results suggest that in β-proteobacteria, substitutions in NtrX alleviated crosstalk with PhoR while in α-proteobacteria substitutions in PhoR prevented crosstalk with NtrX. Although the substitutions are different, the net result in both cases was an insulation of the Ntr and Pho systems.

Global Optimization of Signaling Fidelity

Our results with the Pho and Ntr signaling pathways indicate that the avoidance of crosstalk following gene duplication is a major selective pressure that drives the accumulation of adaptive substitutions in the specificity-determining residues of two-component signaling proteins. More generally, this model
predicts that the specificity residues of two-component signaling proteins in extant organisms should be sufficiently different from, or orthogonal to, one another to prevent crosstalk. To test this prediction, we extracted the six major specificity residues from each of the 22 canonical histidine kinases encoded in the E. coli K12 genome (Figure 5A). Pairwise comparisons indicated that kinases typically had no more than three identities with every other kinase at these six specificity sites, often with nonconservative differences at the remaining sites. One notable exception is NarX and NarQ, which contain two identities and four conservative differences. However, these kinases, which likely arose through gene duplication, each phosphorylate the response regulators NarL and NarP in vitro and likely in vivo, and hence represent a case of physiologically beneficial cross-regulation (Noriega et al., 2010). Aside from these two kinases, there is a general pattern of orthogonality between specificity residues in the system-wide set of E. coli histidine kinases. This orthogonality is further reflected by a lack of information in a sequence logo built from the specificity residues of the 22 E. coli histidine kinases (Figure 5B), particularly in comparison to the sequence logos built from orthogonal histidine kinases (Figures 1B and 3C). A similar pattern of orthogonality was evident in the specificity residues of the 20 canonical histidine kinases in C. crescentus, as well as the specificity residues of the response regulators from both E. coli and C. crescentus (Figure S4). These observations, in combination with our detailed investigation of the Ntr and Pho proteins across phylogenies, suggest that the avoidance of crosstalk is a pervasive and significant selective pressure driving the system-wide insulation of two-component signaling pathways, and consequently, that in extant organisms, two-component systems are largely insulated from one another (Figure 5C).

**DISCUSSION**

Signaling protein families, in both prokaryotes and eukaryotes, frequently expand through gene duplication (Ohno, 1970; Pires-daSilva and Sommer, 2003). The retention of the duplicated genes often requires mutations that insulate them from one another, allowing each to transmit signals without inducing crosstalk. This divergence process may be additionally constrained by a need to avoid crosstalk with other, existing members of the same protein family. For two-component pathways, the duplication-divergence process can be conceptually framed by considering the sequence space defined by the specificity-determining residues of histidine kinases (Figures 5 and 6). For each kinase, these residues dictate the substrates it can phosphorylate, with different kinases recognizing largely nonoverlapping, or orthogonal, sets of substrates. Gene duplication leads initially to a complete overlap and requires that one or both of the duplicates accumulate changes in its specificity residues, thus separating them in sequence space (Figure 6).

The mutational path taken by a given kinase may cause it to infringe on the sequence space occupied by another kinase, as was likely the case with the Ntr and Pho systems in α-proteobacteria. Such overlap then necessitates additional mutations to achieve a system-wide optimization of specificity. Our results indicate that such optimization and the avoidance of crosstalk are important selective pressures influencing two-component

**Figure 5. Extant Two-Component Signaling Pathways Are Insulated from Each Other at the Level of Phosphotransfer**

(A) The six primary specificity residues are shown for each of the 22 canonical E. coli histidine kinases. Hybrid histidine kinases and the noncanonical kinases DcuS and CheA were omitted. The histidine kinases are separated into groups by color based on the family of their cognate response regulator: pink, OmpR/winged helix-turn-helix; green, NtrC/AAA+ and Fis domains; blue, NarL/GerE helix-turn-helix; brown, LytR. For specificity residues from E. coli response regulators and C. crescentus histidine kinases and response regulators, see Figure S4.

(B) Sequence logo for the specificity residues in (A).

(C) A qualitative two-dimensional representation of the distribution of E. coli histidine kinases in the sequence space defined by the six primary specificity-determining residues. Each oval represents the set of response regulators recognized by a histidine kinase given its specificity residues. Spheres are colored using the same scheme as in (A). With the exception of NarQ and NarX (see text), the spheres are nonoverlapping, indicating a lack of crosstalk in vivo and in vitro. Kinases were placed relative to one another based roughly on their ability to phosphorylate the cognate regulators of other histidine kinases after extended incubation times in vitro (Skerker et al., 2005; Yamamoto et al., 2005). For example, CpxA shows a strong preference for phosphotransfer to its cognate regulator CpxR but will phosphorylate the cognate regulators of EnvZ and RstB after extended periods of time.

<table>
<thead>
<tr>
<th>Kinase</th>
<th>Specificity Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>CusS</td>
<td>TNTQEI</td>
</tr>
<tr>
<td>YedV</td>
<td>NAGQQV</td>
</tr>
<tr>
<td>QseC</td>
<td>TAVGEV</td>
</tr>
<tr>
<td>CpxA</td>
<td>TRLGAL</td>
</tr>
<tr>
<td>BasS</td>
<td>AGLHEL</td>
</tr>
<tr>
<td>BaeS</td>
<td>AVGEAA</td>
</tr>
<tr>
<td>PhoQ</td>
<td>AVSTRS</td>
</tr>
<tr>
<td>RstB</td>
<td>VRYREM</td>
</tr>
<tr>
<td>EnvZ</td>
<td>TRLAEM</td>
</tr>
<tr>
<td>PhoR</td>
<td>TVGYEM</td>
</tr>
<tr>
<td>KdpD</td>
<td>TVQGE1</td>
</tr>
<tr>
<td>CreC</td>
<td>AAGAEI</td>
</tr>
<tr>
<td>YfhK</td>
<td>ASEGEL</td>
</tr>
<tr>
<td>ZraS</td>
<td>SSGLKY</td>
</tr>
<tr>
<td>NtrB</td>
<td>GGGQQL</td>
</tr>
<tr>
<td>AtoS</td>
<td>TACYQI</td>
</tr>
<tr>
<td>DpiB</td>
<td>STGQ1</td>
</tr>
<tr>
<td>NarX</td>
<td>LSRCMV</td>
</tr>
<tr>
<td>NarQ</td>
<td>LSRILT</td>
</tr>
<tr>
<td>UhpB</td>
<td>ITRTAG</td>
</tr>
<tr>
<td>YehU</td>
<td>NTAVRR</td>
</tr>
<tr>
<td>YpdA</td>
<td>NASSRL</td>
</tr>
</tbody>
</table>
systems and they can drive the divergent evolution of orthologous proteins.

How did the NtrY-NtrX pathway arise if crosstalk is detrimental? Although we cannot infer the order of events with complete certainty, a plausible scenario is that the NtrY-NtrX pathway arose during growth in phosphate-replete conditions where it provided a selective advantage, as suggested by the slow growth of a ΔntrX strain in these conditions. Subsequent growth in phosphate-limited conditions would then select for strains that had accumulated mutations eliminating crosstalk between PhoR and NtrX. We showed that such insulation can occur with only one or two point mutations, supporting the plausibility of this scenario in an ancestral α-proteobacterium. Interestingly, the β-proteobacteria likely followed a different mutational path to avoiding crosstalk, accumulating substitutions in NtrX rather than PhoR. The difference between the two clades of proteobacteria may reflect the inherent stochasticity of mutations and selection. Alternatively, the growth conditions or genomic context of the ancestral organisms in which gene duplication occurred may have deterministically influenced selection.

In sum, we propose that the evolution of two-component signaling genes is characterized by long periods of stasis with specificity-determining residues subject to strong purifying selection to ensure robust phosphotransfer from kinase to regulator. Gene duplication, or lateral transfer events, can disrupt this stasis, requiring a global reoptimization of existing signaling proteins to accommodate the new pathway. The specificity residues are thus likely subject to bursts of diversifying selection; however, these residues would not necessarily exhibit commonly used signatures of diversifying selection such as large dN/dS values. Instead, our work emphasizes that a molecular-level understanding of protein evolution and the identification of adaptive mutations ultimately demands an integration of sequence analysis with focused biochemical and genetic characterizations.

Our approach and findings are relevant beyond two-component signaling as paralogous signaling protein families are found throughout biology. In fact, most organisms use a remarkably small number of types of signaling protein to carry out their diverse information-processing tasks. In all cases, duplication and divergence is a primary means by which new pathways are created and, consequently, issues of specificity and the fidelity of information transfer are critical. While eukaryotes sometimes rely on tissue-specific expression of paralogous genes or spatial mechanisms like scaffolds to enforce specificity, many common signaling proteins and domains, such as PDZ, SH3, SH2, and bZIP proteins (Hou et al., 2009; Liu et al., 2011; Newman and Keating, 2003; Stiffler et al., 2007; Tonikian et al., 2008; Zarrinpar et al., 2003), rely on molecular recognition and a relatively small set of specificity-determining residues. Hence, our observation that pathway insulation in bacteria can be achieved with a limited number of mutations may help to explain how organisms in all domains of life have exploited gene duplication to expand and diversify their signaling repertoires.

**EXPERIMENTAL PROCEDURES**

**Identification of Orthologs and Construction of Gene Trees**

A modified version of reciprocal best blast hits was used to identify orthologous proteins. For *E. coli* PhoR, the DHp domain was used as a query in BLAST searches against fully sequenced bacterial genomes in GenBank (September 2009). The top ten hits from each genome were then subjected to reciprocal BLAST searches against the *E. coli* MG1655 genome. If only the top hit identified *E. coli* PhoR as the best match, it was called as a PhoR ortholog. If multiple hits identified *E. coli* PhoR as the best match, the top hit was called as an ortholog and additional hits were evaluated as follows. If an additional hit had an E-value within 10^{-5} of the top hit and was closer to the top hit than to the fifth hit (which generally had an E-value reflecting the overall paralogous relationship of histidine kinases), we also called it as an ortholog of *E. coli* PhoR and examined the next hit similarly. For genomes with more than one hit called as an ortholog, duplications were inferred and each hit deemed a member of the PhoR orthogroup. A similar procedure was followed to identify orthologs of PhoB, NtrB, NtrC, and NtrY using as query sequences the *E. coli* PhoB receiver domain, the *C. crescentus* NtrB and NtrY DHp domains, or the *C. crescentus* NtrC and NtrX receiver domains.
Orthologous sequences were aligned using ClustalX (Chenna et al., 2003). Sequence logos for the specificity residues, extracted from the aligned sequences, were built using WebLogo (Crooks et al., 2004). To help correct for phylogenetic biases in genome sequencing efforts, sequences were filtered to ensure that no two sequences were more than 95% identical.

PhoR Dhp domains and PhoB receiver domains were extracted using HMMER with models for a His_KA domain or REC domain, respectively (Wistrand and Sonnhammer, 2005), and used to build gene trees through the PHYLIP package (Felsenstein, 1989) using the neighbor-joining algorithm provided. The tree was rooted using Bacillus subtilis PhoR as the outgroup. Reported bootstrap values are out of 1,000.

For genome-wide analyses of specificity residues, only canonical histidine kinases were included. Canonical kinases were defined as those containing the PFAM HisKA domain and no REC domain.

### Growth Conditions and Strain Construction

*C. crescentus* cells were grown at 30°C in PYE, M2G (10 mM phosphate), M5G (50 μM phosphate), or MBG (same as M5G but with 5 μM phosphate), supplemented when necessary with oxytetracycline (1 μg/ml) or kanamycin (25 μg/ml), or 0.3% xylose.

#### Protein Purification and Phosphotransfer Assays

Expression, protein purification, and phosphotransfer experiments were carried out as described previously (Skerker et al., 2005). Phosphotransfer profiles against all ntrX regulators except for CC1741 (50 μM phosphate), or M5G. Transductions were performed using 4°C (Ely, 1991).

Strains used are listed in Table S3. To construct ML1934, PhoR(TV), a region from nucleotide position 30 (relative to the PhoR start codon) to position 168 was amplified using CB15N genomic DNA as template and the primers PhoR_int_upstream_for and PhoR_int_upstream_rev, and a region from 147 to 1416 was amplified using pENTR-CC_PhoR(TV) as template and the primers PhoR_int_for and PhoR_int_rev. Primer sequences are listed in Table S4. The two amplicons were then fused using SOE-PCR and ligated into pNPTTS138, which had been cut with EcoRV and phosphatased using SAP, to create pNPTTS-CC_PhoR(TV). This plasmid was used for allelic replacement in CB15N following procedures described previously (Skerker et al., 2005). Integrants were tested for kanamycin sensitivity, sucrose resistance, and sequence-verified using primers listed in Table S4. To create ML1935, PhoR(TV)/ntrX, a tetracycline-marked ntrX deletion was transduced into ML1934; transductants were verified by PCR.

Strains used for competition assays (ML1936-ML1945) contained either the coding region for CFP or YFP driven by the xyI promoter. xyI was amplified from CB15N genomic DNA using primers xyI_for and xyI_rev, digested with KpnI and AgeI and ligated into pXCFPN-4 and pXYFPN-4 using the same restriction sites, producing pXCFPN-4:PxyI and pXYFPN-4:PxyF-yfp. The inserts were then amplified using primers xyI_xf0_for and xyI_xf0_rev, digested with HindIII and EcoRI, and ligated into pNPTTS138 digested with the same enzymes. These vectors, pNPTTS138:PxyI and pNPTTS138:PxyF-yfp, were then integrated into the chromosomes of CB15N, ΔntrX, PhoR(TV), and PhoR(TV)/ntrX through transformation and selection on kanamycin followed by counterselection on sucrose, leading to markerless integrations of CFP or YFP at the native xyI locus of each strain.

Expression vectors were built by moving pENTR clones into destination vectors using the Gateway LR reaction (Invitrogen), and then transformed into BL21 E. coli for expression and purification. All site-directed mutagenesis was done on pENTR clones using primers listed in Table S4 and sequence verified.

#### Growth and Competitive Fitness Assays

Cultures were grown overnight in M2G. Cultures were then diluted to OD600 ~0.025 and resuspended in either M2G or M5G. Samples were taken every hour and growth rates calculated 8–14 hr postdilution to ensure phosphate limitation of cells grown in M5G. For more severe phosphate limitation, growth curves were repeated in MBG, which contains 10-fold less phosphate. For these experiments, cultures were grown overnight in M5G and resuspended at an OD600 ~0.07 in MBG.

For competitive fitness assays, cultures were grown overnight in M2G, resuspended in either M2GX or M5GX to an OD600 of 0.05, and mixed 1:1 with a competitor strain in 10 ml of media in a 150 ml flask. After 9 hr, a sample was taken and fixed using paraformaldehyde. Cells were then diluted to OD600 ~0.01 in 10 ml of M2GX or M5GX. After 15 hr, another sample was taken and cells diluted to OD600 ~0.05. After 9 hr, another sample was taken and cells diluted to OD600 ~0.01. This growth and dilution process was repeated until 104 total hours had elapsed. Cultures typically remained below OD600 ~0.85 at all times. Cells from each sample collected were immobilized on 1.5% agarose pads made with PBS, and imaged using a Zeiss Axiovert 200 microscope with a 100x objective. Multiple fields of CFP, YFP, and phase images were taken for each sample. Roughly 500 cells were counted for each time point using a custom MATLAB script with counts checked manually using ImageJ. Competition experiments were done once with wild-type expressing cfp and mutant expressing yfp, repeated with fluorescent proteins swapped, and results averaged.

### Microarray Analysis

Cultures were grown to mid-log phase in M2G and either RNA was harvested or cells were washed, resuspended in M5G, and grown for 11 hr in phosphate-limited conditions before RNA was harvested. RNA was extracted, labeled, and hybridized to custom-designed 8x15K Agilent expression arrays as described previously (Gora et al., 2010). NtrX-dependent genes were defined as those genes exhibiting at least a 4-fold decrease in expression in the ntrX strain compared to wild-type C. crescentus in M2G. Complete array data are provided in Table S2 and deposited in GEO.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.05.033.

### ACKNOWLEDGMENTS

We thank O. Ashenberg for help with bioinformatics, Y.E. Chen for help with strain construction, and A. Podgornaia, A. Keating, O. Ashenberg, and K. Foster for helpful comments on the manuscript. Sequence analyses were performed on a computer cluster supported by NSF grant 0821391. M.T.L. is an Early Career Scientist of the Howard Hughes Medical Institute. This work was supported by an NSF graduate fellowship to E.J.C. and an NSF CAREER award to M.T.L.

Received: March 19, 2012

Revised: April 23, 2012

Accepted: May 18, 2012

Published: July 5, 2012

### REFERENCES


